Atherosclerosis 258 (2017) 1–7

Contents lists available at ScienceDirect

Atherosclerosis

journal homepage: www.elsevier.com/locate/atherosclerosis

Genetic risk analysis of coronary artery disease in Pakistani subjects using a genetic risk score of 21 variants

Saleem Ullah Shahid [a, *, 1], Shabana [a, 1], Jackie A. Cooper [b], Katherine E. Beaney [b], Kawah Li [b], Abdul Rehman [a], Steve E. Humphries [b]

[a] Department of Microbiology and Molecular Genetics, University of the Punjab, Lahore, 54590, Pakistan
[b] Centre for Cardiovascular Genetics, British Heart Foundation Laboratories, University College London, London, WC1E6JF, UK

ARTICLE INFO

Article history:
Received 19 July 2016
Received in revised form
17 January 2017
Accepted 19 January 2017
Available online 22 January 2017

Keywords:
Coronary artery disease
Genetic risk score
Conventional risk factors

ABSTRACT

Background and aims: Conventional coronary artery disease (CAD) risk factors like age, gender, blood lipids, hypertension and smoking have been the basis of CAD risk prediction algorithms, but provide only modest discrimination. Genetic risk score (GRS) may provide improved discrimination over and above conventional risk factors. Here we analyzed the genetic risk of CAD in subjects from Pakistan, using a GRS of 21 variants in 18 genes and examined whether the GRS is associated with blood lipid levels.

Methods: 625 (405 cases and 220 controls) subjects were genotyped for variants, *NOS3* rs1799983, *SMAD3* rs17228212, *APOB* rs1042031, *LPA* rs3798220, *LPA* rs10455872, *SORT1* rs646776, *APOE* rs429358, *GLUL* rs10911021, *FTO* rs9939609, *MIA3* rs17465637, *CDKN2A*rs10757274, *DAB2IP* rs7025486, *CXCL12* rs1746048, *ACE* rs4341, *APOA5* rs662799, *CETP* rs708272, *MRAS* rs9818870, *LPL* rs328, *LPL* rs1801177, *PCSK9* rs11591147 and *APOE* rs7412 by TaqMan and KASPar allele discrimination techniques.

Results: Individually, the single SNPs were not associated with CAD except *APOB* rs1042031 and *FTO* rs993969 ($p = 0.01$ and 0.009 respectively). However, the combined GRS of 21 SNPs was significantly higher in cases than controls (19.37 ± 2.56 *vs.* 18.47 ± 2.45, $p = 2.9 \times 10^{-5}$), and compared to the bottom quintile, CAD risk in the top quintile of the GRS was 2.96 (95% CI 1.71–5.13). Atherogenic blood lipids showed significant positive association with GRS.

Conclusions: The GRS was quantitatively associated with CAD risk and showed association with blood lipid levels, suggesting that the mechanism of these variants is likely to be, in part at least, through creating an atherogenic lipid profile in subjects carrying high numbers of risk alleles.

1. Introduction

Coronary artery disease is a chronic disorder progressing silently and usually has established to an advance stage by the time symptoms start appearing. Despite all measures, CAD remains the single largest killer worldwide. In high income countries, the CAD mortality rate has declined since 1980 and has shifted to an older age group, whereas, middle and low income countries bear three quarters of the global CAD burden. South Asians are at a greater risk and the prevalence is 50%–300% higher than rest of the world [1]. The prevalence of CAD is even higher in Pakistan [2] with more than 30% of the population above 45 years of age being affected by the disease [3]. The disease burden has almost doubled in urban Karachi since 1970 [4]. According to latest WHO reports, cardiovascular diseases (CVD) are among the biggest non-communicable killers in Pakistan and CAD represents a major type of CVD (http://www.who.int/countries/pak/en/).

CAD is a multifactorial disorder and arises from an interaction between environmental and genetic factors. The identifiable environmental risk factors have been identified in about 80% of CAD cases [5]. Most of the CAD risk factors are modifiable therefore, to target life style changes or for drug intervention, those who are at most risk of developing disease should be identified earlier. The conventional CAD risk factors (CRFs) like age, gender, blood lipids, smoking, blood pressure and diabetes have been the basis of CAD risk prediction algorithms developed by many consortia. These risk prediction algorithms include the Framingham risk score [6], the Prospective Cardiovascular Munster Heart Study (PROCAM) [7], the

* Corresponding author.
E-mail address: saleemullahshahid@hotmail.com (S.U. Shahid).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.atherosclerosis.2017.01.024
0021-9150/

Systematic Coronary Risk Evaluation (SCORE) system [8], the Reynolds risk score [9] and QRISK2 [10]. These CRF algorithms calculate 10 years CAD risk and the individuals are then classified according to their risk category. The high risk individuals qualify for the preventive treatment (statin). Until recently, the cut-off for statin treatment has been set at a 20% 10 years CAD risk [11]. Lower cut-off value for statin treatment has been proposed in both UK and USA (10% and 7.5% respectively). The use of CAD risk prediction scores has increased the average life time of CAD patients by three years in USA [12]. However, the risk assessment using CRFs provides only modest discrimination and does not fully explain the underlying risk [13]. These risk scores lack accuracy and may overestimate the risk in low risk subjects or underestimate the risk in subjects at high risk [6,14]. Almost 15–20% cases who developed CAD in their later lives were underestimated [15] with most of the cases occurring in intermediate and low risk subjects [16,17].

The variability in disease susceptibility in individuals exposed to similar environmental factors and having almost similar CRFs may be attributed to the genetic variations [18]. Genetic testing may improve discrimination over and above the CRFs alone. A family history of early heart disease has long been a known risk factor and heritability of CAD has been estimated to be more than 40% [19]. Historically, the genetic risk of CAD was assessed through the presence of the disease in the proband's relatives and the genetic component was described as heritability estimate. Then the 'candidate gene' approach was used, where the common variants in the genes regulating biochemical pathways of disease pathogenesis were determined [20]. Since 2007, additional genes associated with CAD have been identified through Genome Wide Association studies (GWAS) [21]. Single nucleotide polymorphisms (SNPs) can be used as markers of genetic variability. The SNPs associated with CAD are common in general population with a minimal to moderate relative risk. Most of them are located in non-coding DNA region implying that they may be influencing by regulating the expression of upstream or downstream genes. Another striking feature of CAD risk SNPs is that most of them operate independently of known CAD risk factors. This indicates that many unknown pathways involved in the development of CAD still need to be explored [22]. However, the risk associated with a single SNP is modest, because of the low effect sizes of common variants, and therefore a large number of SNPs needs to be genotyped for the genetic analysis of CAD like complex disease.

A GRS of a disease is calculated by summing up the number of risk alleles at all the loci included in the genetic risk analysis of that disease. The GRS is a multi-locus profile used to transpose the discoveries from candidate gene studies and GWASs into population health tools [23,24]. A GRS summarizes the effect of multiple variants in a quantitative manner and hence is superior over the predictive power of a single SNP. The use of GRS information in CAD risk prediction can bridge up the genomic research with more applied clinical practice. Different researchers have used varying number and types of loci for inclusion in CAD genetic risk, the number ranging from less than 10 to more than 100 [25–29].

The majority of genetic studies and GWAS have been conducted on European/Caucasian people. It remained a routine practice to transpose the results obtained from such studies conducted in developed countries to the rest of the world, but there remains an immense requirement to extend genetic studies to other ethnicities also. The allele frequencies of many common variants vary widely between ethnicities. For example, the association of the 9p21 region with CAD has not been replicated in African Americans [30,31]. Similarly, the linkage disequilibrium and effect size of common variants may vary across different ethnicities. Moreover, a genetic marker may not be associated with a trait in all ethnicities and in such cases the applicability is limited to only those populations where the genotype to phenotype association is clearly seen [32]. The Pakistani population, like other Asian countries is underrepresented in international genetic studies like HAP MAP or 1000 genomes project. To date, the genetic architecture of CAD has not been properly evaluated for this population. A preliminary report of the use of a 19 SNPs GRS in CAD risk analysis in the Pakistani subjects has been published [33], but the study was underpowered to detect the same effect as observed in Europeans (308 cases and 130 controls). In the current study, we included two additional SNPs to construct a CAD GRS and increased the sample size including 405 cases and 220 controls. We hypothesized that to predict CAD risk in the Pakistani subjects, a GRS of 21 SNPs will be superior over single SNPs having small effect size and modest association.

2. Materials and methods

The study comprised of 405 diagnosed CAD cases and 220 healthy controls. The criteria for the selection of study subjects has been described previously [34]. The CAD cases were recruited from tertiary care hospitals in Lahore during February 2012 to June 2013. The selected subjects had suffered from a non-fatal myocardial infarction with diagnosis made by the consultant cardiologist based on the reports of ECG, cardiac echo, angiography, troponine T/I and clinical history. Only those CAD cases were selected which were recently diagnosed and had not started lipid lowering or antihypertensive drugs therapy. The controls were apparently healthy subjects, not having any history of an early CAD in their family. It was taken care that cases and controls represented all the socio-economic groups. Subjects with obesity (BMI>26 kg/m^2 for Asian populations as described previously [35]) were excluded from the study to reduce the possible confouders but not those with Type 2 diabetes because the number of CAD subjects with type 2 diabetes was high and the sample size would have become too small to have adequate power. All participants gave a written informed consent. The study was approved by the ethics committee, University of the Punjab, Lahore and all the procedures were in compliance with the Helsinki declaration.

2.1. Genotyping

The DNA was extracted from whole blood leucocytes using Wizard genomic DNA purification Kit (Promega, USA). The DNA samples were quantified using nanodrop (ND-8000, USA). The concentration of DNA samples was standardized to 1.25ng/ul. The genotyping was carried out in especially designed 384 well plates (Micro Amp). The DNA samples were arrayed into plates by an automated robotic liquid handling system (Biomerk-FX, Beckman Couter). Two high throughput florescence based allele discrimination techniques, TaqMan and KASPar, were used for genotyping the SNPs. The details of genotyping techniques have been given somewhere else [36]. The information on SNPs included in the study is provided in Supplementary Table 1.

The SNPs *NOS3* (rs1799983), *SMAD3* (rs17228212), *APOB* (rs1042031), *LPA* (rs3798220), *LPA* (rs10455872), *SORT1* (rs646776), *APOE* (rs429358), *GLUL* (rs10911021) and *FTO* (rs9939609) were genotyped by TaqMan technique using qPCR master mix (KAPA Biosystems, USA). The SNPs *MIA3* (rs17465637), *CDKN2A* (rs10757274), *DAB2IP* (rs7025486), *CXCL12* (rs1746048), *ACE* (rs4341), *APOA5* (rs662799), *CETP* (rs708272), *MRAS* (rs9818870) *LPL* (rs328), *LPL* (rs1801177), *PCSK9* (rs11591147) and *APOE* (rs7412) were genotyped by KASPar technique with touchdown thermal cycler programme. The SNP *LPL* (rs1801177) was monomorphic in this population, but data from this SNP is shown for completeness. The list of primers and probes used for TaqMan and KASPar are